

# An approximation scheme for the generalized geometric minimum spanning tree problem with grid clustering

Citation for published version (APA):

Grigoriev, A., & Feremans, C. (2004). *An approximation scheme for the generalized geometric minimum spanning tree problem with grid clustering*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 036 <https://doi.org/10.26481/umamet.2004036>

## Document status and date:

Published: 01/01/2004

## DOI:

[10.26481/umamet.2004036](https://doi.org/10.26481/umamet.2004036)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# An Approximation Scheme for the Generalized Geometric Minimum Spanning Tree Problem with Grid Clustering

CORINNE FEREMANS    ALEXANDER GRIGORIEV \*

September 9, 2004

## Abstract

This paper is concerned with a special case of the Generalized Minimum Spanning Tree Problem. The Generalized Minimum Spanning Tree Problem is defined on an undirected graph, where the vertex set is partitioned into clusters, and non-negative costs are associated with the edges. The problem is to find a tree of minimum cost containing exactly one vertex in each cluster. We consider a geometric case of the problem where the graph is complete, all vertices are situated in the plane, and Euclidean distance defines the edge cost. We prove that the problem admits PTAS if restricted to grid clustering.

**Keywords:** Generalized Minimum Spanning Tree, Approximations, Grid Clustering.

## 1 Introduction

**Problem Statement.** Consider the *Generalized Minimum Spanning Tree Problem (GMST)*. Let  $G = (V, E)$  be an undirected graph with vertex set  $V$  and edge set  $E \subseteq V \times V$ . Given edge costs  $c_e \in \mathbb{R}^+, e \in E$ , and a partition of the vertex set  $V$  into non-intersecting clusters  $V_i, i \in K$ , the problem is to find a tree of minimum cost containing exactly one vertex in each cluster. This problem is a generalization of the well known *Minimal Spanning Tree Problem (MST)* where every vertex of the graph is a cluster itself.

This article considers a *geometric* version of GMST with *grid clustering*. We assume that the graph is complete, and all  $n = |V|$  vertices are situated inside the  $k \times \ell$  planar integer grid. The edge costs correspond to the Euclidean distance between the points in the plane. We assume that all vertices belonging to the same grid cell form a cluster, and there is at least one vertex in each grid cell. Without loss of generality we assume that there are no vertices of the graph on the borders of the grid cells.

---

\*c.feremans, a.grigoriev@ke.unimaas.nl. Department of Quantitative Economy, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Fax: +31-(0)43-3884874

**Applications.** Applications for GMST are encountered in telecommunications where local subnetworks must be interconnected by a global network containing a gateway from each subnetwork. For this inter-networking, a vertex has to be chosen in each local network as a hub and the hub vertices must be connected via transmission links such as optical fiber (see e.g. Myung, Lee and Tcha [5]).

**Literature Overview.** The GMST was introduced by Myung, Lee and Tcha [5], where the authors show that the problem is strongly  $\mathcal{NP}$ -hard and there is no polynomial time algorithm solving the problem with bounded worst-case ratio, unless  $\mathcal{P} = \mathcal{NP}$ . Notice that there is a huge complexity jump from polynomially solvable MST, see e.g. Papadimitriou and Steiglitz [6], to constant inapproximable GMST. Woeginger, [9], shows that the geometric version of GMST is also strongly  $\mathcal{NP}$ -hard. The reductions for the geometric GMST is based on the reduction for the geometric *Steiner Minimal Trees* presented by Garey, Graham and Johnson [4]. Exact search methods and heuristics have been developed for GMST (see Feremans, Labbé and Laporte [3] for a survey and see Duin, Volgenant and Voß[1] for the latest reference about computational results). Some positive approximation results for GMST are described in Pop, Kern and Still [8] providing a polynomial time approximation algorithm with worst-case ratio bounded by  $2\rho$  if the cluster size is bounded by  $\rho$ .

Geometric GMST instances with grid clustering and with the same number of vertices in each cluster were widely used to test methods implemented to solve GMST, see e.g. Feremans [2], Pop [7]. This class of GMST has never been theoretically studied in itself. To our knowledge, there is nothing known on the approximability and complexity of this problem.

**Our results and organization of the paper.** In this paper we close the lack of knowledge on the approximability of the problem although its complexity status remains open. The main result of the article is a polynomial time approximation scheme (PTAS) for the geometric GMST with grid clustering. This PTAS is based on a dynamic programming algorithm solving the problem in polynomial time when  $k$  or  $\ell$  is bounded from above by a constant. Existence of PTAS shows that the geometric version of GMST with grid clustering is much easier to solve than GMST itself. In this sense, an extensive usage of instances of the geometric GMST with grid clustering to test sophisticated algorithms solving GMST is not adequate.

The remainder of the article is organized as follows. In Section 2 we introduce a dynamic programming algorithm that solves the problem in polynomial time when  $k$  or  $\ell$  is bounded. In Section 3, we introduce a PTAS based on a dynamic programming approach for the geometric GMST with grid clustering. In Section 4 extensions of presented algorithms are discussed.

**Notation.** Since we have a very structured clustering, indices of the clusters are defined by indices of the grid cells:  $V_{r,t}$ ,  $r \in \{1, \dots, k\}$ ,  $t \in \{1, \dots, \ell\}$ . Index  $r$  correspond to the  $y$ -axe of the plane and index  $t$  corresponds to the  $x$ -axe.

Without loss of generality we assume that  $k \leq \ell$ . We denote the union of cells in row  $r \in \{1, \dots, k\}$  by  $R_r$  and the union of cells in column  $t \in \{1, \dots, \ell\}$  by  $C_t$ . Finally, let the cardinality of the maximal cluster be  $\rho = \max_{r,t} |V_{r,t}|$ .

An example of the problem instance is given in Figure 1 for  $k = 3, \ell = 5, \rho = 4$ , and  $n = 30$ .

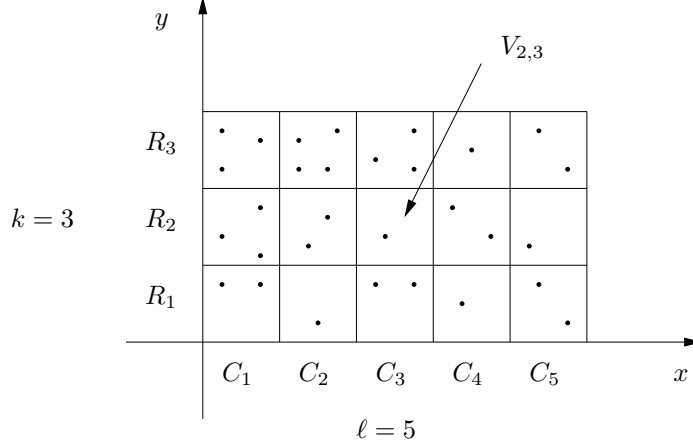


Figure 1: Problem instance for  $k = 3, \ell = 5, \rho = 4$ , and  $n = 30$ .

## 2 Dynamic Programming Algorithm

In this section we construct a dynamic programming algorithm that solves the geometric GMST with grid clustering in polynomial time if  $k$  is bounded by a constant.

The following lemma is a key to our dynamic programming algorithm. Actually, the lemma shows that an optimal solution of a geometric GMST with grid clustering does not contain long edges. Dynamic programming can then step by step construct a tree using only local information.

**Lemma 2.1** *An optimal solution of a geometric GMST with grid clustering does not contain edges with projection on  $x$ -axes or  $y$ -axes longer or equal to 3.*

**Proof.** Consider an arbitrary generalized minimum spanning tree  $T^{OPT}$ . Assume by contradiction that in  $T^{OPT}$  there exists a long edge  $e = (v, u)$  crossing two or more vertical grid lines and having  $x$ -projection longer or equal to 3. Consider the set  $Q_e$  of the grid cells having common points with  $e$ . Since  $e$  is a continuous line in the plane, the grid cells in  $Q_e$  are connected. Deletion of  $e$  from  $T^{OPT}$  disconnects the tree into two components. Moreover,  $v$  and  $u$  belong to different components. Therefore, by continuity there are two adjacent cells  $V^1 \in Q_e$  and  $V^2 \in Q_e$  and two vertices  $v' \in V^1$  and  $u' \in V^2$  belonging to  $T^{OPT}$  such that  $v'$  and  $u'$  belong to different components. Consider a generalized spanning tree  $T'$  obtained from  $T^{OPT}$  by deletion of edge  $e$  and introduction of new edge  $e' = (v', u')$  (see Figure 2).

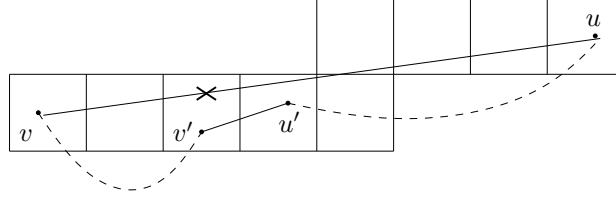


Figure 2: No long jump lemma.

Since  $v'$  and  $u'$  are two vertices from adjacent cells, we derive that  $c(e') \leq 2\sqrt{2}$ . By assumption  $c(e) \geq 3$ . Hence,

$$c(T') = c(T^{OPT}) - c(e) + c(e') \leq c(T^{OPT}) - 3 + 2\sqrt{2} < c(T^{OPT}),$$

that contradicts the optimality of  $T^{OPT}$ .  $\blacksquare$

Starting from the left and moving to the right, in step  $t$  of the algorithm,  $t \in \{1, 2, \dots, \ell - 3\}$ , we consider four consecutive columns  $C_t$ ,  $C_{t+1}$ ,  $C_{t+2}$  and  $C_{t+3}$  of the grid. Let  $S_t$  be a subset of  $V$  containing exactly one vertex from each cell in  $C_{t+1}$ ,  $C_{t+2}$  and  $C_{t+3}$ . Clearly,  $|S_t| = 3k$  for any  $t \in \{1, \dots, \ell - 3\}$ . Let  $[M]_{3k \times 3k}$  be a symmetric zero-one *transitivity* matrix, where by transitivity we mean that if  $M_{p,q} = 1$  and  $M_{q,s} = 1$  then  $M_{p,s} = 1$ . Let us denote the set of all such matrices by  $\mathcal{M}$ . A matrix  $M \in \mathcal{M}$  represents the connectivity of vertices in  $S_t$  by a path in the partial solution induced by the first  $t + 3$  columns of the grid. More precisely, let  $f(S_t, M)$ , be a generalized *minimum* spanning forest on the first  $t + 3$  columns provided that (1) in any cell of the first  $t$  columns there is a vertex connected (not necessarily adjacent) to some vertex in  $S_t$ ; (2) two vertices  $v$  and  $u$  in  $S_t$  are connected if and only if  $M_{v,u} = 1$ .

By Lemma 2.1 any forest  $f(S_t, M)$  can be obtained as a forest  $f(S_{t-1}, M')$  extended by a subset  $T_t$  of edges on the vertex set  $S_{t-1} \cup S_t$ . Moreover, given a connectivity matrix  $M' \in \mathcal{M}$  for set  $S_{t-1}$ , a connectivity matrix  $M \in \mathcal{M}$  for set  $S_t$ , and any subset  $T_t$  of edges on  $S_{t-1} \cup S_t$ , we can easily verify whether connectivity  $M$  on  $S_t$  can be obtained by introducing edges  $T_t$  to a graph with connectivity  $M'$  on  $S_{t-1}$ . If the combination of  $M'$  and  $T_t$  is consistent with  $M$  we write  $M' \oplus T_t = M$ . Now, the cost of forest  $f(S_t, M)$  can be found by the following recurrent formula

$$c(f(S_t, M)) = \min_{S_{t-1}, M', T_t: M' \oplus T_t = M} (c(f(S_{t-1}, M')) + \sum_{e \in T_t} c_e).$$

At step  $\ell - 3$  the algorithm stops and output the solution

$$c(f^*) = \min_{S_{\ell-3}} c(f(S_{\ell-3}, \mathbf{1})),$$

where  $\mathbf{1}$  is the unit matrix from  $\mathcal{M}$ .

Let us estimate the running time of the dynamic programming algorithm. At each step of the algorithm we calculate the values  $c(f(S_t, M))$  for all possible subsets  $S_t$  and all matrices  $M \in \mathcal{M}$ . The number of subsets  $S_t$  is  $O(\rho^{3k})$  and the number of matrices  $M$  is  $O(2^{9k^2})$ . Given a subset  $S_t$  and a matrix  $M$ , to

calculate the value  $c(f(S_t, M))$  we again enumerate all possible sets  $S_{t-1}$  and all possible matrices  $M'$ . In addition we enumerate all edge sets  $T_t$  on  $4k$  vertices. The number of such sets is  $O(2^{16k^2})$ . Given  $S_{t-1}$ ,  $M'$ , and  $T_t$ , the calculation of  $c(f(S_{t-1}, M')) + \sum_{e \in T_t} c_e$  takes  $O(k^2)$  time. Thus, the total running time of the dynamic programming algorithm is  $O(\ell \rho^{6k} 2^{34k^2} k^2)$  which is polynomial if  $k$  is fixed.

Notice that the algorithm can be speeded up if we take into account that the edge set  $T_t$  must be a forest and that in edge set  $T_t$  there are no edges with  $y$ -projection longer or equal to 3.

We summarize the outcome of this section in the following theorem.

**Theorem 2.2** *A dynamic programming algorithm solves the geometric GMST with grid clustering in time  $O(\ell \rho^{6k} 2^{34k^2} k^2)$ .* ■

### 3 Polynomial Time Approximation Scheme

In this section, we describe an algorithm that provides a polynomial time approximation scheme for the geometric GMST with grid clustering. The idea of the algorithm is as follows.

Given an accuracy parameter  $\varepsilon > 0$ , let us split the grid into  $\Delta = \varepsilon(k-1)/54\sqrt{5}$  equal *slices*: rows  $R_1, \dots, R_{k/\Delta}$  form the first slice, rows  $R_{(k/\Delta)+1}, \dots, R_{2k/\Delta}$  form the second one, etc. Notice that the number of rows in a slice is bounded by  $O(1/\varepsilon)$ , and therefore we can find the generalized minimum spanning tree in every slice in polynomial time by dynamic programming described above. Then we connect the generalized minimum spanning trees in slices by picking up the shortest edge between the trees in each pair of neighboring slices. Clearly, the resulting tree  $T^{APPX}$  is a generalized spanning tree. It remains to show that the cost  $c(T^{APPX})$  is only factor of  $(1+\varepsilon)$  far from the cost of generalized minimum spanning tree  $c(T^{OPT})$ .

**Theorem 3.1**

$$\frac{c(T^{APPX}) - c(T^{OPT})}{c(T^{OPT})} \leq \varepsilon.$$

**Proof.** By definition  $T^{APPX}$  consists of  $\Delta$  generalized minimum spanning trees  $T_i, i \in \{1, \dots, \Delta\}$ , on the slices and  $\Delta - 1$  link edges between these trees. The link edges have length at most  $\sqrt{5}$  (taking the diagonals crossing two cells). Therefore

$$c(T^{APPX}) \leq \sqrt{5}(\Delta - 1) + \sum_{i=1}^{\Delta} c(T_i). \quad (1)$$

Consider a generalized minimum spanning tree  $T^{OPT}$  and an arbitrary slice  $i \in \{1, \dots, \Delta\}$  of the grid. Let  $F_i$  be the forest induced by  $T^{OPT}$  on the slice  $i$ , so

$$c(T^{OPT}) \geq \sum_{i=1}^{\Delta} c(F_i). \quad (2)$$

Let us remind that by Lemma 2.1 in  $T^{OPT}$  there are no edges with  $y$ -projection longer or equal to 3. It implies that in introducing new edges in  $F_i$  to provide the connectivity of all vertices in the upper three rows and in the bottom three rows in slice  $i$ , we automatically obtain the connectivity of all vertices in  $F_i$ , i.e. after adding these edges, forest  $F_i$  becomes a connected tree. Since a tree can connect all vertices in the upper three rows of the slice (the same for the bottom three rows), we need at most  $2(3\ell - 1)$  edges of length at most  $\sqrt{5}$  to transform  $F_i$  into a tree, therefore for any  $i \in \{1, \dots, \Delta\}$

$$c(F_i) + \sqrt{5}(6\ell - 2) \geq c(T_i). \quad (3)$$

Combining inequalities (1), (2), and (3), we derive

$$c(T^{APPX}) - c(T^{OPT}) \leq \sqrt{5}(\Delta - 1) + \sum_{i=1}^{\Delta} c(T_i) - \sum_{i=1}^{\Delta} c(F_i) \leq 6\sqrt{5}\ell\Delta. \quad (4)$$

Now, to complete the performance guarantee analysis for PTAS we need a lower bound on the optimal solution. Consider a partition of the  $k \times \ell$  grid into  $3 \times 3$  subgrids (without loss of generality we assume that  $k$  and  $\ell$  are divisible by 3). The connection of the center cell of a subgrid to its boundary takes at least length 1. Therefore, to connect the subgrids together we need at least length  $k\ell/9 - 1$ .

Now, we are ready to estimate the relative ratio between the approximate and the optimal solutions of the problem

$$\frac{c(T^{APPX}) - c(T^{OPT})}{c(T^{OPT})} \leq \frac{54\sqrt{5}\ell\Delta}{k\ell - 9} \leq \frac{54\sqrt{5}\Delta}{k - 1} = \varepsilon, \quad (5)$$

that completes the proof. We assume  $\ell \geq 9$  else, an optimal solution is computed by complete enumeration.  $\blacksquare$

## 4 Extensions of the presented results.

The settings for the generalized GMST with grid clustering were chosen in the sake of simplicity of explanations. The reader can easily verify that the results of the paper can be straightforwardly generalized or adopted for

- arbitrary size of the grid cells;
- the case where all vertices of the graph are situated in the  $k \times \ell$  grid, the cells are the clusters, empty clusters are allowed, all non-empty clusters are connected, and the number of non-empty clusters is of order  $\Omega(k\ell)$ ;
- the case with any regular partition of the plane into the clusters, for instance, clustering by proper triangles or clustering by proper hexagons;
- other problems like generalized version of *minimum  $k$ -vertex connected subgraph* with grid clustering and low connectivity requirement.

## References

- [1] C. W. Duin, A. Volgenant, S. Voß. *Solving group Steiner problems as Steiner problems*. European Journal of Operational Research **154** (2004) 323–329.
- [2] C. Feremans. *Generalized Spanning Trees and Extensions*. Ph.D. dissertation. Université Libre de Bruxelles (2001).
- [3] C. Feremans, M. Labbé, G. Laporte. *Generalized Network Design Problems*. European Journal of Operational Research. **148** (2003) 1–13.
- [4] M.R.Garey, R.L.Graham, D.S.Johnson. *The Complexity of Computing Steiner Minimal Trees*. SIAM Journal of Applied Mathematics **32** (1977) 835–859.
- [5] Y.S. Myung, C.H. Lee, D.W. Tcha. *On the Generalized Minimum Spanning Tree Problem*. Networks **26** (1995) 231–241.
- [6] C.H. Papadimitriou, K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications, New York (1998).
- [7] P.C. Pop. *The Generalized Minimum Spanning Tree Problem*. Ph.D. Thesis. University of Twente (2002).
- [8] P.C. Pop, W. Kern, G.J. Still. *An Approximation Algorithm for the Generalized Minimum Spanning Tree Problem with Bounded Cluster Size*. Working Paper, Departement of Operations Research and Mathematical Programming, University of Twente, <http://www.math.utwente.nl/dos/ormp/preprints.htm> (2001).
- [9] G.Woeginger. Private Communications (2004).